

基於辭典詞彙釋義之多階層釋義關聯程度計量—— 以「目」字部為例

A Measurement of Multi-Level Semantic Relations among Mandarin Lexemes with Radical *mu4*: A Study based on Dictionary Explanations

趙逢毅*、鍾曉芳⁺

F. Y. August Chao, Siaw-Fong Chung

摘要

本研究使用辭典中詞語與釋義的關係，透過計算機進行詞語的多階層釋義關聯 (Multi-Level Semantic Relation) 量化計算，以比較字與詞組之間的釋義關聯程度。本研究從臺灣「教育部重編國語辭典修訂本」之中，取出屬「目」字部之所有字/詞共計 4549 個字與詞彙進行直接試驗，以避免人為語意辨別模糊與使用不同語料庫定義不同的缺點。經過不同的試驗說明多階層釋義關聯的特色及使用方式，其中包括將屬「目」字部所有的字與「目」進行多階層釋義關聯計算，表現出漢字意符的特色。最後並與常用的 MI Value 及 t-score 比較多階層釋義關聯之異同。

關鍵字：釋義關聯，多階層釋義關聯網路，辭典，語料庫，「目」字部

¹ 本論文出處於中國蘇州舉辦之「第十一屆漢語詞彙語義學研討會(CLSW2010)」增修版本。

* 國立政治大學資訊管理研究所，台灣台北市文山區指南路二段 64 號

Department of Management Information Systems, National Chengchi University

E-mail: fychao.tw@gmail.com

⁺ 國立政治大學英國語文學系

Department of English, National Chengchi University

E-mail: sfchung@nccu.edu.tw

Abstract

In this study, we utilize a quantitative method measuring the Multi-Level Semantic Relations based on 4549 Mandarin lexemes containing the radical *mu4* (目). The research is carried out by first extracting all dictionary definitions for all lexemes containing this radical. Then, we consider the different layers of definitions (e.g., the definitions of the keywords in a definition) and measure whether two different *mu4* (目) lexemes are related in meanings. It was found that both width (the number of lexemes covered) and depth (the number of levels to be calculated) contribute to the measurement of semantic relatedness. Some seemingly unrelated *mu4* (目) lexemes are found related when the depth of definitions increases. The study also compares two sets of results - one based on MI value and the other based on t-score. Our findings show that our measurement based on multi-level semantic relations produces better results than MI value does, as a collocation measurement like MI value is less suitable for analyzing semantically related dictionary entries.

Keywords: Definition relation, Multi-Level Semantic Relation, Dictionary, Corpus, Mandarin radical *mu4* (目)

1. 前言

「辭典」乃依據詞彙體系及一定的編輯體例蒐集詞、詞組、短語等資料，加以解釋以備查索、參考的工具書。在編輯辭典時，除了要考量到查詢者所擁有的先備知識(Prior Knowledge)以撰寫其能理解的釋義之外，還要使用簡單的釋義文字說明，才能使查詢者瞭解該字所俱備的義涵。中文辭典的排版則是透過具有意符表徵的部首，將所有漢字與詞語聯繫並形成知識架構(周亞民、黃居仁，2005)。黃居仁(2005)亦認為詞是「語言中表達意義的最小獨立單位」，在辭典中則是透過不同的詞彙組織成爲釋義文字以說明查詢的辭彙。因此從辭典的釋義文字、詞組與查詢字(詞)之間的釋義關聯，可以發現釋義說明所用的單一詞語會包括被解釋字(詞)的部份或完全的涵義，以及對該字(詞)認知和意義延伸的比喻涵義，並且使用屬於知識層級上較爲通俗的語句或概念撰寫釋義詞語，從而讓具有一般知識水準的大眾都能很快理解釋義的說明。臺灣的「教育部重編國語辭典修訂本」即是在提供各界人士及中小學生檢索、查閱的目的之下而編修的一本實體紙本工具書，因此編輯釋義內容均使用大眾能理解的文字及簡單的詞語說明辭典之中的字詞。

在分析字(詞)之間的釋義關聯方法中，除了透過人爲語感對詞彙進行詞意訊息上的討論外(黃居仁等 2005, 2008)，便是使用已經建立好的詞彙工具進行語意的探索，如：知網(HowNet)(Dong & Dong, 2003)、英文的 Wordnet (Miller, 1995)及中研院的中文詞彙網路(Chinese WordNet, CWN)(Huang, Chang, & Lee, 2004)。雖然現有語料工具定義詞彙語意十分詳細，卻無法如辭典釋義般詳述詞彙語意的內容。本論文透過使用 Chung, Chen, 與 Chao (2009) 所建立的隱喻關聯程度計算，直接計算中文辭典之中的釋義文字，同時比較

字(詞)之間的釋義關聯(dictionary definition relationship)。除了建立釋義關聯網路與同義字彙集群之外，更進一步擴展此關聯計算以揭露中文詞彙之間，在相同部首之中深層的釋義關聯。

2. 相關研究

中文字本身就存在很強的義形(plyph)與概念(concepts)的聯結關係。在編撰辭典時，參與編輯人員亦會仔細研究該字要表達的意義與所歸屬的部首之間概念上是否恰當(周亞民、黃居仁，2005)。鄭文泉(2004)更從符號學的角度指出「漢字藉由方位隱喻足以使人們領略它的肖像的符號涵意」。其它學者則從不同的觀點討論此一課題，如黎傳緒(2004)對“相”的解析，討論「本義」與「引義」之間的發展關係。另一方面，祝清(2009)從詞類上著手，探討漢語獨立詞類—動名詞，提出動名詞實為語法的隱喻，是詞類的去範疇化與再範疇化的結果。

中文辭典則是將不同的辭句，透過文字部首架構起來匯集成為辭典。辭典釋義的編撰會因操作視角策略與適用的領域不盡相同而造成釋義不同，但都是透過文字紀錄下適合於當代的定義與闡釋的語料庫(羅益民，2007)。從語料庫的觀點而言，透過部首分類的字或詞組應可在其所包含的義涵之中尋找到概念上的聯繫。黃居仁等(2005, 2008)則以人工的方式，從知識概念的層級上建立中文部首與文字之間的關係，透過物質(formal)、組成(constitutive)、功用(telic)、事件(participating)、參與者(participator)、描述狀態(descriptive)與產生(agentive)等七類衍生面向，探討漢字知識表達的層面與目、耳、口、鼻、舌等五官類漢字意符(漢字構字要件)的語意關聯。為能確保詞彙語意的品質與一致性，黃居仁、蔡柏生等(2003)則認為參與編撰人員需對詞義判準有一致性的準則，避免不同人對某一詞彙在語境中，依據不同的直覺而有不同的區分方式。

另一方面，近十多年來語料工具發展對詞彙語意探討，與詞彙語意之間關係的規範有很大的幫助。常見的語料工具有：Wordnet (Miller, 1995)，一個由普林斯頓大學所發展出的英文語料工具，其中定義了詞彙的涵義、上/下位關係詞、部份/全體關係詞與同義詞；Hownet (Dong & Dong, 2003)，以常識(common sense)定義中文字彙的義原，並討論義原之間的關聯所建立的詞彙知識庫；不同於 Hownet 的義原關聯建立原則，中研院的中文詞彙網路(Chinese WordNet, CWN)(Huang, Chang, & Lee, 2004)使用中文詞義(sense)區分資料，並且將中文詞義與英文詞義(Wordnet)建立關聯。透過使用前述不同的中文/英文語料工具，高照明(2007)同時整合不同辭典中文詞彙語意訊息，建立詞彙關係擷取系統。然而前述中文語料工具所涵蓋的字與詞彙相較於辭典字彙仍有不足外，語料工具所建立的知識領域亦有所局限，而且使用英文語料工具還有語言平行對譯等問題。

不同於前述，本研究直接探討字與詞彙間釋義的描述關聯。延伸自 Chung *et al.* (2009)對知識本體與語料庫之中的字詞，以字詞共同出現的知識概念出現比例，決定隱喻關聯的強度，藉此建立辭典字彙之間的釋義關聯網絡。

3. 研究方法

本研究收集提供給廣泛使用者查詢的「教育部重編國語辭典修訂本」資料(網路版)，透過統計詞彙之間共同出現的釋義字彙，以了解詞彙之間多階層釋義關聯程度，並說明此關聯程度所表達的意思之間的概念關聯。首先呈現資料收集原則與比較釋義關聯計算，並說明辭典資料透過此關聯度而表現出不同程度數值的釋義關聯內容。接著詳述延伸自隱喻關聯計算原則的多階層釋義關聯，並驗證黃居仁(2005, 2008)部首與中文字之間的意符關聯。最後則將比較多階層釋義關聯與互見信息值 (Mutual Information Value)、t-score 間的不同。

3.1 資料收集與實作關聯計算

在中文辭典中，部首的分類原則在民國初年已經過許多人的改制，目前較通用的檢字規則是教育部頒訂的「部首檢字法」(改自林語堂「上下形檢字法」)，將中文字所表達的意符分類到不同部首中。因此本文在資料收集的過程之中，為求所收集到的辭典釋義能將含括特定意符的知識概念都包含在詞彙的釋義之中，因此在收集的字詞選擇上，參考薛榕婷(2003)對部首與文化之間的研究，選定依「目」字部為主，並收集所有包括屬「目」字部首的字詞共計 4549 字(字數 115 與詞彙數 4434)，其中亦包括異體字(如瞅、眀、眈等同為「目」字的異體字與變形部首，如：眾與曩)。在取得釋義之後，將資料透過中研院所開發的中文斷詞系統(<http://ckipsvr.iis.sinica.edu.tw/>)以獲得釋義文字之中各詞語的詞類。在取得字詞類的釋義後，本研究保留知識概念涵義較大的動詞詞類('VA', 'VAC', 'VB', 'VC', 'Vi', 'Vt', 'VCL', 'VD', 'VE', 'VF', 'VG', 'VH', 'VHC', 'VI', 'VJ', 'VK', 'VL', 'V_2')與名詞詞類('Na', 'Nb', 'Nc', 'Ncc', 'Nd', 'N')進行隱喻關聯的計算。此外亦將標記後句中含有分號(COLONCATEGORY)的句子刪除，因在「教育部重編國語辭典修訂本」的釋義中，使用到分號的句子為來源或例句的說明。接著以「瞄」、「瞄準」二字為例說明處理過程：

「瞄」為「注視」；

「瞄準」為「用眼睛注視目標，使發射、投射的動作準確。」

此二組字詞之間因包含「注視」而產生了釋義關聯。所取得的釋義在經過斷詞系統處理及詞類過濾後，保留下的釋義分別為：

「瞄」為「注視(VC)」；

「瞄準」為「眼睛(Na)注視(VC)目標(Na)使(VL)發射(VC)投射(VC)動作(Na)」

從上述的資料可知，「瞄」與「瞄準」共同使用的釋義詞為「注視」。本研究的釋義關聯計算，參考知識概念隱喻計算(Chung *et. al.*, 2009)，以共同出現的二度隱喻詞百分比表示 (“Percentage of co-appearance of 2nd degree relations”, CoAP)。其中令 x 為來源詞語的字詞，其過濾後符合的字詞集合為 X ，而 y 為目標詞語，其過濾後符合的字詞集合為 Y ； $X \cap Y$ 則為兩組字詞共同所使用的詞組， x 、 y 個別的二度隱喻詞百分比為：

$$CoAP(x) = \frac{X \cap Y}{X}, \quad (1)$$

在 X 所有釋義中與 Y 共同出現的字詞的比例，同理

$$CoAP(y) = \frac{X \cap Y}{Y}, \quad (2)$$

即 Y 所有釋義中與 X 共同出現的字詞的比例。

依前例，「瞄」字的釋義為：“注視(VC)”、「瞄準」的釋義僅保留：“眼睛(Na) 注視(VC) 目標(Na) 使(VL) 發射(VC) 投射(VC) 動作(Na)”，則

$$CoAP(\text{瞄}) = \frac{(\text{注視})}{(\text{注視})} = 1, \text{ 同理}$$

$$CoAP(\text{瞄準}) = \frac{(\text{注視})}{(\text{眼睛、注視、目標、使、發射、投射、動作})} = \frac{1}{7}$$

「瞄」字對「瞄準」的釋義關聯大於「瞄準」對「瞄」的釋義關聯。意即「瞄」所含的概念(即共有的概念詞彙, "注視VC")能被「瞄準」所包括, 而「瞄準」的概念卻無法透過「瞄」的概念完整詮釋。最後再透過網路分析工具 Pajek (<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>) 進行釋義關聯網路的繪製, 如圖 1。圖中表示「瞄」與「瞄準」(深色方型, 辭典詞彙)在釋義之中都有使用到“注視”(淺色菱形, 辭典釋義中的釋義關聯媒介文字)。由此可得知, 「瞄」與「瞄準」在辭典釋義之中都會使用“注視”來描述兩者涵義。

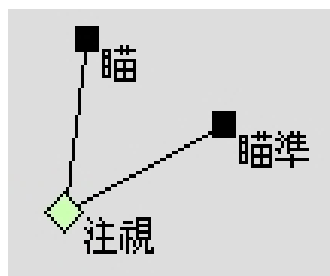


圖 1. 「瞄準」與「瞄」之語意媒介網路 (菱形為釋義關聯媒介, 方型為辭典詞語)

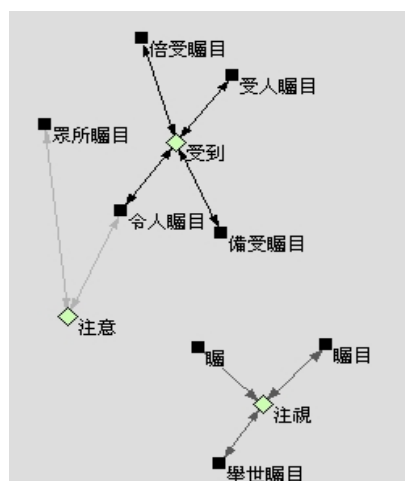


圖 2. 「矚」及包涵「矚」字之詞語間交互語意媒介網路

同理, 計算「矚」與其相關字詞(包括「令人矚目」、「倍受矚目」、「備受矚目」、「受人矚目」、「引人矚目」、「眾所矚目」、「矚望」、「矚目」、「學世矚目」、「遠矚」、「駭矚」、「高瞻遠矚」、「麗矚」)等 14 個詞語, 將其中產生語意關聯的 8 個詞語繪成語意媒介網路為圖 2。從圖 2 可知「矚」及使用到「矚」字的 8 個詞語之中, 依釋義的動詞及名詞的使用情況可以區分為兩個群, 即圖 2 中的半下方群集與“注視”有釋義關聯

的「矚」群，及屬於「注意」-「受到」釋義的群。爲了能更清楚地了解集群中的義涵，我們試著從釋義中了解中介節點的涵義如下：「受到」爲單一動詞；「注意」的釋義爲：關注留意；「注視」的釋義爲：集中視線，凝目而望。因此「注視」群相較於「注意」-「受到」群保有實際上的視線、凝望等釋義關聯，而「注意」-「受到」釋義的辭典詞彙群則有接受及注意的釋義關聯。由此可知，應用知識概念隱喻計算於辭典釋義之中，再藉由詞彙語意媒介網路的集群具體呈現，結合詞彙釋義的深入了解，我們可以更清楚地區別辭典如何詮釋字彙的涵義，也更明白近義詞彙之間的不同形式。

進一步分析字詞時，由於釋義關聯媒介文字的比例會隨辭典詞彙的比例增加，使得釋義關聯無法將高度關聯的詞彙表現出來。因此參考 Chung *et. al.* (2009) 的隱喻關聯計算，將兩詞語相互之間共同詞彙所擁有之詞彙比例進行平均數計算，以表示來源及目標釋義關聯相互之間的釋義關聯程度 (Definition Relation Degree, DRD)：

$$DRD_{xy} = 2 \frac{CoAP(x)CoAP(y)}{CoAP(x) + CoAP(y)}, 0 \leq DRD_{xy} \leq 1 \quad (3)$$

承前例，「瞄」與「瞄準」兩詞之間的釋義關聯程度即爲：

$$DRD_{\text{瞄, 瞄準}} = 2 \frac{CoAP(\text{瞄})CoAP(\text{瞄準})}{CoAP(\text{瞄}) + CoAP(\text{瞄準})} = 2 \frac{1(\frac{1}{7})}{1 + \frac{1}{7}} = 0.25 = DRD_{\text{瞄準, 瞄}}$$

從「瞄」與「瞄準」的釋義之中可知，雖然兩者共同使用「注視」爲釋義詞，但因爲「瞄準」的釋義詞使用較多名詞與動詞，即「瞄準」的概念較「瞄」的概念複雜，從而降低了兩詞彙間的釋義關聯程度。

表1. 「矚」及包涵「矚」字之詞間交互釋義關聯程度表

詞(一)	詞(二)	詞(一)對詞(二) 之 CoAP	詞(二)對詞(一) 之 CoAP	詞(一)與詞(二) 的 DRD
矚	矚目	1	1	1
備受矚目	受人矚目	1	0.5	0.67
倍受矚目	備受矚目	0.5	1	0.67
倍受矚目	受人矚目	0.5	0.5	0.5
令人矚目	備受矚目	0.2	1	0.3
令人矚目	受人矚目	0.2	0.5	0.29
倍受矚目	令人矚目	0.5	0.2	0.29
眾所矚目	令矚目	0.3	0.2	0.25
舉世矚目	矚目	0.125	1	0.22
矚	舉世矚目	1	0.125	0.22

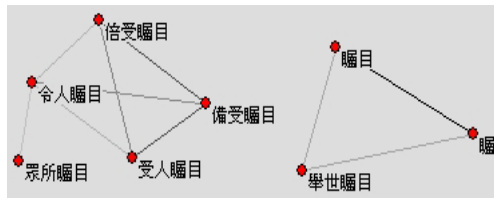


圖3. 「矚」及包括「矚」字之釋義關聯網路
(顏色深淺表DRD 數值，顏色深表數值高)

依上述原則，計算「矚」及包括「矚」之辭典詞彙(見表 1)，並透過釋義關聯程度(DRD)值計算，簡化成單純釋義關聯網路，並去除釋義關聯網路中間的媒介釋義字詞(見圖 3)。在表 3 中，詞(一)與詞(二)為來源與目的辭典條目，透過「詞(一)對詞(二)之 CoAP」與「詞(二)對詞(一)之 CoAP」分別計算兩者間共用釋義媒介詞的占有比率後，再取兩 CoAP 去方向性的平均值為釋義關聯值(DRD)。

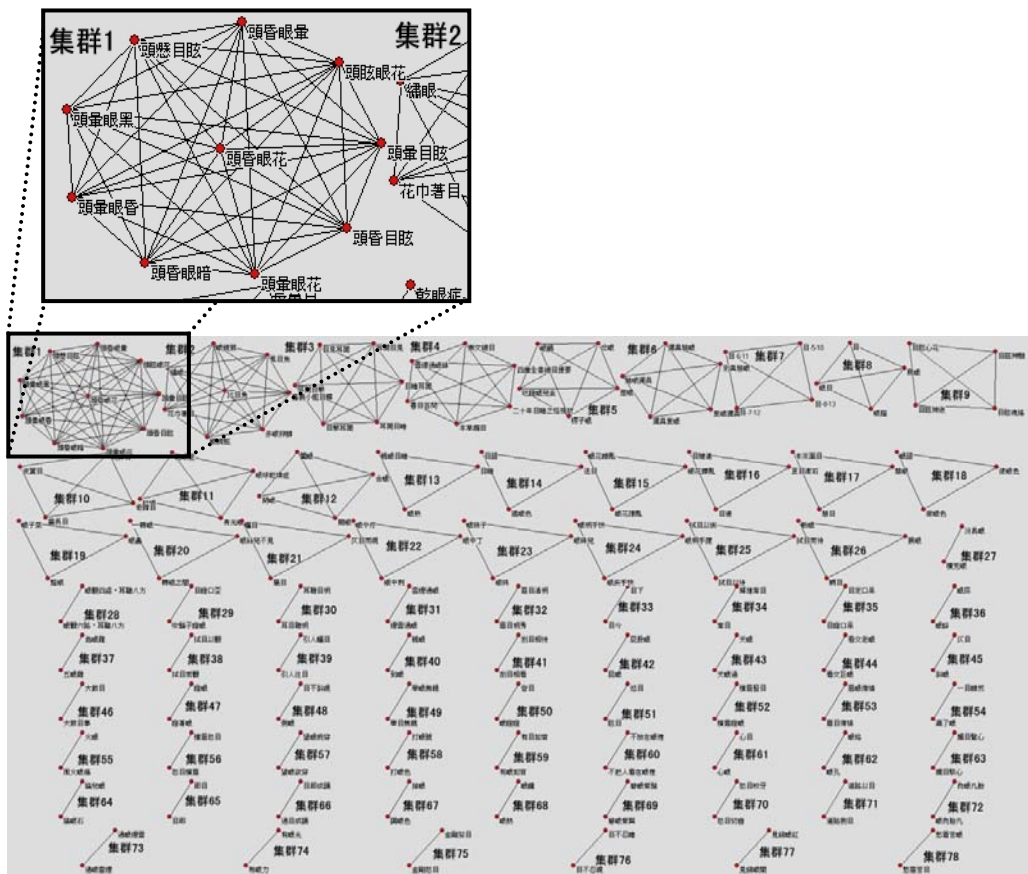


圖4. 「目」、「眼」及包括二字的高釋義關聯集群

釋義關聯網路可以藉由群集的結果，瞭解辭典詞彙之間大體上的釋義關聯情況及相關的程度。以「矚」及包括「矚」之辭典詞彙為例，圖 3 左側「倍受矚目」、「備受矚目」及「受人矚目」，與右側的「矚」及「矚目」的釋義關聯程度都較其它辭典詞彙高(表 1 中，DRD 值高於 0.5)。從釋義文字中得知(見圖 2)，「眾所矚目」及「令人矚目」的釋義關聯之中還包括了“注意”的文字概念，而在另一個集群之中的「舉世矚目」，因釋義文字多於「矚」與「矚目」，從而使其釋義關聯程度(在圖 3 呈現) DRD 值降低。因此從資料中我們可以說明，「矚」-「矚目」與「倍受矚目」-「備受矚目」-「受人矚目」這兩個群組的釋義(涵義)上有高度的關聯，而「舉世矚目」的釋義較接近「矚」-「矚目」群而非「倍受矚目」-「備受矚目」-「受人矚目」群。

我們更進一步從前述釋義關聯分析「眼」字與「目」字及包括二字的辭典詞彙之釋義關聯，並分析兩兩成對的釋義關聯程度值 DRD 介在 0.9~1 的高釋義關聯部份，最後藉由社會網路分析軟體 Pajek 中的網路圖佈局集群將此關聯呈現視覺化圖形。結果得到上圖 4 共 78 個集群 (詳見表 2)。

表 2. 「目」、「眼」及包括二字的詞語集群關聯媒介文字表，DRD>0.9

集群別	釋義關聯媒介文字	集群別	釋義關聯媒介文字	集群別	釋義關聯媒介文字
集群 1	頭腦、昏沉、視覺、模糊	集群 11	病名	集群 21	注視
集群 2	動物、名	集群 12	閉上、眼睛	集群 22	比喻、痛恨、人
集群 3	親自、看到、聽到	集群 13	看見	集群 23	眼球
集群 4	書名	集群 14	目、示意	集群 24	眼光、銳利、動作、敏捷
集群 5	北平、方言	集群 15	形容、眼睛、昏花、心緒、迷亂	集群 25	擦亮、眼睛、等待
集群 6	具有、特殊、眼光、見解	集群 16	佛教、人名	集群 26	眨眼
集群 7	異體字	集群 17	眼睛	集群 27 之後為兩個辭典詞彙，且為高度釋義關聯，故不再贅述。	
集群 8	眼睛	集群 18	眼睛、示意		
集群 9	形容、見、情景、令、人、驚異	集群 19	植物、名		
集群 10	哺乳綱	集群 20	眼睛、轉		

其中集群 1 為最大，包括了十個辭典詞語(「頭眩眼花」、「頭暈目眩」、「頭昏眼暗」…等等)，且都是相似的詞義；集群 2 則是表示動物的辭典詞彙；集群 3 則是以「聽到」、「看到」為釋義關聯媒介的集群；集群 4 是書名；集群 5 是北平方言；集群 6 是以「特殊」、「眼光」及「見解」所組成的；集群 7 為異體字集；集群 8 即為眼睛…等等。由此分析可知，DRD 值高即為同義詞，因為用以釋義的字詞幾乎相同。在此要特別

說明的是，透過釋義關聯程度 DRD 值所呈現出的同義詞集群，並非由辭典在編撰時特別說明為同義詞組；而是經由釋義之中所擁有共同詞彙的百分比計算而得。因此，獲得集群後必需再進一步了解各個集群之中的釋義媒介文字為何，才能說明該集群的特徵。

在了解高關聯度 DRD 之後，為討論中度釋義關聯值，我們以「目所未睹」與目字所有詞組進行分析，並取出 DRD 值界在 0.3~0.6 的中度關聯 DRD 詞彙，共九個詞彙列於表 3。

表 3. 「目所未睹」與同屬「目」字部之詞組比較 (DRD 值介於 0.3~0.6 間)

詞彙	釋義	釋義過濾後的詞組	與「目所未睹」相同釋義詞組	DRD 值
目所未睹	從來沒見過，形容極為罕見。	見,形容,罕見	n/a	n/a
目下有臥蠶	形容眼瞼浮腫，下瞼有臥蠶樣。多見於腎炎病人。	形容,眼瞼,浮腫,有,臥蠶樣,見,於,腎炎,病人	形容 見	0.33
目連	人名。見「目犍連」條。	人名,見,目犍	見	0.33
目眩神迷	形容所見情景令人驚異。亦作「目眩神搖」。	形容,見,情景,令,人,驚異,作,目眩神搖	形容 見	0.36
目斷飛鴻	極目遠望，直至飛雁不見。常形容離別的悲悽之情。	遠望,直至,飛雁,見,形容,離別,悲悽,情	形容 見	0.36
目眩魂搖	形容所見情景令人驚異。亦作「目眩神搖」。	形容,見,情景,令,人,驚異,作,目眩神搖	形容 見	0.36
目瞪口呆	形容非常生氣的樣子。	形容,生氣,樣子	形容	0.33
目眩神馳	形容所見情景令人驚異。亦作「目眩神搖」。	形容,見,情景,令,人,驚異,作,目眩神搖	形容 見	0.36
目下十行	形容看書看得快。見「一目十行」條。	形容,看書,看,快,見,一目十行	形容 見	0.33
目視雲霄	形容眼界高闊。	形容,眼界,高闊	形容	0.33

從表 3 可知，九個詞彙分別透過釋義媒介文字“形容”、“見”與辭典條目「目所未睹」產生釋義關聯；但由於與「目所未睹」所共用的詞組太少，而且九個詞彙的釋義又多，從而降低了 DRD 值。使用較多的釋義文字說明詞彙，即表示該詞彙的概念較為複雜。從表 3 的資料中可看到，與「目所未睹」產生釋義關聯的詞彙為「形容」與「見」，雖然此二釋義詞並非詮釋該辭條的主要用意，但可知此中度釋義關聯的主要集群概念都為比喻用詞彙。

同理在我們分析 DRD 值更小的如「眼」與「眼鏡蛇」兩個詞彙(表 4)。因「眼鏡蛇」是屬於表示特定動物的下位詞，從而在辭典編撰中即會使用包含較多上位詞概念的釋義文字以闡述此一詞彙的概念，即該辭彙所涵蓋的概念多且複雜，從而降低了兩個詞彙之間的 DRD 值。

表4. 「眼」與「眼鏡蛇」之詞組比較

詞彙	釋義	釋義過濾後的詞組	釋義中與「眼」相同詞組	DRD 值
眼	目，動物的視覺器官。	目 Na 動物 Na 視覺 Na 器官 Na	n/a	n/a
眼鏡蛇	動物名。爬蟲類有鱗目。長四、五尺，頸部有一對有白邊黑心的環狀斑紋，形如眼睛。背褐色，腹青白色，毒牙有溝，可注毒液，怒則頸膨大，昂首直立，晝伏夜出。俗稱為「飯匙倩」。	動物 N 名 N 爬蟲類 N 有 Vt 鱗目 N 長 Vt 頸部 N 有 Vt 有 Vt 白邊 N 黑心 Vi 環狀 N 斑紋 N 形 N 如 P 眼睛 N 背 Vt 褐色 N 腹 N 青白色 N 毒牙 N 有 Vt 溝 N 注 Vt 毒液 N 怒 Vi 頸 N 膨大 Vi 昂首 Vi 直立 Vi 晝伏夜出 Vi 俗稱為 Vt 飯匙倩 N	動物	0.05

綜合上述分析，直接使用釋義關聯程度(DRD)所得的值，因受詞彙所使用的釋義文字長短或涵蓋的知識概念多寡，而有不同的影響。此外若所使用的釋義文字沒有完全的對應，則無法納入相同詞組的計算。最後釋義關聯計算單元是以字與辭彙所含的「知識概念」，因所有過濾後的辭彙字詞是相等權重，因此在計算上則不需為個別知識概念加權。反觀釋義文字在編撰時，會因該詞彙概念內容特殊而特別偏重在幾個詞上。因為 DRD 的釋義關聯計算上，是將取自釋義的詞組視為包括部分或全部的知識概念單元進行計算，從而無法完整表達個別釋義詞組所含蓋知識概念的權重。

3.2 多階層釋義關聯

前述已討論過，應用基於辭典釋義的釋義關聯程度值 DRD 計算時，會因釋義中單詞的權重與解釋文句長短而影響 DRD 值。為了改善釋義關聯計算使其更加精確，本研究提出多階層釋義關聯(Multi-level Definition Relation Degree, MDRD)。黃居仁指出(2005)漢語的詞是知識表達的最基本單位，因此在釋義之中所使用的詞組也可以再透過辭典的釋義，擴充共同使用的字詞而併入計算的權重中。本研究所建立的辭典語料是以部首屬「目」的字詞，因此進行釋義文字擴充時會以現有的辭典語料進行擴充。

令 X^1 為條目 x 所有符合過濾條件的辭典釋義詞彙； X^2 為 X^1 釋義詞彙再經過辭典釋義文字擴充並符合過濾條件詞彙；同理， Y^1 為 Y^1 符合的釋義詞彙，再經辭典釋義擴充後且符合過濾條件詞彙集合。而 $X^{(1+2)} \cap Y^{(1+2)}$ 則分別為 X^1 、 X^2 與 Y^1 、 Y^2 共同擁有的釋義

文字的集合。而 $CoAP_{x,y}^2(x) = \frac{(X^1 + X^2) \cap (Y^1 + Y^2)}{(X^1 + X^2)}$ ，表在 x 所有擴充釋義中與 y 共同使

用的擴充釋義字詞的比例，則第 n 階層釋義關聯的一般式即可寫成：

$$CoAP_{x,y}^n(x) = \frac{\sum_{i=n} X^i \cap \sum_{i=n} Y^i}{\sum_{i=n} Y^i} \quad (4)$$

$$DRD_{x,y}^n = 2 \frac{CoAP_{x,y}^n(x) CoAP_{x,y}^n(y)}{CoAP_{x,y}^n(x) + CoAP_{x,y}^n(y)}, \quad 0 \leq DRD_{x,y}^n \leq 1 \quad (5)$$

當釋義關聯計算加入擴充的釋義文字之後，釋義的字詞片段會依所搜集來自辭典語料而增加該字詞的權重與增加計算的詞語廣度，進而產生且增加兩者釋義語意上的關聯程度。以「目的」與「瞄準」為例(參考次頁表 5)，使用 DRD 計算「目的」與「瞄準」之第一階共同釋義字百分比與釋義關聯值：

$$CoAP(\text{瞄準}) = \frac{1}{3}, \quad CoAP(\text{目的}) = \frac{1}{9}, \quad DRD_{\text{目的, 瞄準}} = 2 \frac{\left(\frac{1}{3}\right) \times \left(\frac{1}{9}\right)}{\left(\frac{1}{3}\right) + \left(\frac{1}{9}\right)} = 0.167$$

在計算第二階層釋義關聯時，二字的釋義都經過部首屬「目」的字與辭彙進行擴充後並納入計算，則兩者之間的第二階層釋義關聯 $SRD_{\text{目的, 瞄準}}^2$ 為：

$$CoAP^2(\text{瞄準}) = \frac{27}{40} = 0.675, \quad CoAP^2(\text{目的}) = \frac{27}{28} = 0.964$$

$$DRD_{\text{目的, 瞄準}}^2 = 2 \frac{(0.675) \times (0.964)}{(0.675) + (0.964)} = 0.794$$

從上述的計算可知，透過將第一階層釋義文字內的「目標」一詞的釋義擴充後，「目的」與「瞄準」兩詞之間的釋義關聯提高到 0.794。其中，共同擁有的釋義文字，在第一階段僅只有“目標”一個媒介詞。但在經過二階段擴充之後，“達到”與“注視”兩字會因釋義有共同交集而使權重增加。媒介文字“達到”在「目的」的第一階段的釋義文字與第二階段透過「目標」擴充後，釋義文字產生加權效果；同理“注視”在「瞄準」的釋義之中亦產生了加權效果。此外，兩者之間的釋義關聯，也因為「目標」一詞的擴充而加強了彼此的釋義關聯強度。在此特別說明，此擴充之釋義計算僅涵蓋現有辭典語料，即「目」字部所有字辭。

雖然共有的詞組數都為 27，但「目的」的共有詞組表中，較「瞄準」多計算了分別落於第一階釋義(「目的」的直接釋義)與第二階釋義詞語(擴充自「目的」釋義詞組中的“目標”釋義詞組)的“達到”一詞。而相反的，「瞄準」則較「目的」多計算了一組“注視”的釋義詞語(分屬於「瞄準」的直接釋義與「目標」的擴充釋義)。由此可進一步推論，「目的」相較於「瞄準」有地理位置轉移的概念義涵；「瞄準」相較於「目的」，則著重在視覺專注的狀態或動作。而從 $CoAP^2$ 的數值上看來，「目的」的概念義涵較能表達「瞄準」概念，但「瞄準」的概念涵義則無法表達「目的」的義涵，因「瞄準」一詞的義涵較複雜。

表5. 「目的」與「瞄準」之多階層釋義表 (共同出現的釋義已特別標示)

詞彙	釋義	釋義過濾後的詞組	共有的詞組數
第一階段釋義			
目的	想要達到的目標。如：「人生以服務為目的。」	想要, <u>達到</u> , <u>目標</u>	1
瞄準	用眼睛注視目標，使發射、投射的動作準確。	用, 眼睛, <u>注視</u> , <u>目標</u> , 使, 發射, 投射, 動作, 準確	
第二階段釋義			
目標	可為目力的標準或目力能注視的地方。工作或計畫中擬訂要達到的標準。軍事上運用軍隊所望達成的最終目的，或攻擊行動所望殲滅的敵軍或攻占的地區或地點。	為, 目力, 標準, 目力, <u>注視</u> , 地方, 工作, 計畫, 擬訂, <u>達到</u> , 標準, 軍事, 上, 運用, 軍隊, 望, 達成, 目的, 行動, 望, 殲滅, 敵軍, 攻占, 地區, 地點	
眼睛	動物身上觀察外物的視覺器官。	動物, 身, 觀察, 外物, 視覺, 器官	
合併計算之釋義詞組及權重 (字詞後的數字為計次)			
目的	<u>行動 1 攻占 1 計畫 1 運用 1 望 2 地區 1 殲滅 1 敵軍 1 為 1 工作 1 上 1 地方 1</u> <u>注視 1</u> <u>軍事 1</u> <u>達到 2</u> <u>軍隊 1 標準 2 目力 2 目標 1 目的 1 地點 1 擬訂 1 達成 1 想要 1</u>		27
瞄準	<u>行動 1 攻占 1 計畫 1 運用 1 望 2 地區 1 殲滅 1 敵軍 1 為 1 工作 1 上 1 地方 1</u> <u>注視 2</u> <u>軍事 1</u> <u>達到 1</u> <u>軍隊 1 標準 2 目力 2 目標 1 目的 1 地點 1 擬訂 1 達成 1 投射 1 發射 1 身 1 觀察 1 準確 1 視覺 1 眼睛 1 外物 1 器官 1 用 1 使 1 動作 1 動物 1</u>		27

由於此辭典的釋義只限於部首屬「目」的字與辭彙，因此在表5第一級釋義裡僅有「目標」與「眼睛」兩個詞是會被擴充的。其它不屬於部首「目」的字詞不會再經釋義擴充，因此這些不再擴充釋義字詞的權重相對於能擴充的字詞來說權重較低，也就是透過字詞的擴充決定不同釋義詞的權重。另一方面，雖然釋義關聯的計算仍以字形為主的計算比較原則，但中文經過斷字詞之後的字詞組保有釋義文字要傳遞的知識單元，因此再經過許多階層釋義擴充之後，便能透過使用相同的釋義文字而說明兩個辭彙之間的關聯。依據前述的計算方式，我們計算「目的」與「目」、「目的」與「眼」之間的釋義關聯(表6)。在「目的」第一階層的辭典釋義之中並沒有包括任何直接與「眼睛」一詞有關的字詞，但過濾後的詞「目標」經多階層釋義擴充之後則揭露出「眼睛」的概念，並與「目」與「眼」詞產生語意上的關聯。

表6. 「目的」與「目」、「目的」與「眼」第1~4階層釋義關聯值

x, y	$DRD_{x,y}^1$	$DRD_{x,y}^2$	$DRD_{x,y}^3$	$DRD_{x,y}^4$	第1~4階層 共有的媒介釋義字
目的, 目	0	0	0.049	0.298	身 觀察 視覺 眼睛 外物 器官 動物
目的, 眼	0	0	0.044	0.297	身 觀察 視覺 眼睛 外物 器官 動物

接著討論多階層釋義關聯中的階層特性。我們使用相同的語料庫進行擴充多階層的釋義，並計算從第一階到第七階層、第一百階的「眼腦」與包括「眼」字的雙字詞釋義關聯值(圖5)。圖中所設定的第一百階層釋義關聯值，是爲了確保所有釋義的字詞都已經透過語料庫擴充，也就是多階層釋義關聯值在本語料庫之中已經呈穩定狀態(釋義關聯值不會因爲再擴充而改變)。

「眼腦」的釋義爲"眼睛"，但在使用單階層的釋義關聯計算時，會無法與「眼睛」的釋義文字進行概念的計算，從而使 DRD 值爲 0。同樣的情況發生在後續許多雙字詞之中，如「眼圈」、「眼眉」、「眼裡」等 33 個雙字詞。但隨著多階層語意的擴展後，將個別雙字詞與「眼腦」之間共字使用的釋義字詞納入計算後，其 DRD 數值也隨之增加。另一方向，隨著納入計算的擴充釋義字詞的增加，會直接影響釋義字詞占 DRD 權重，DRD 值會突然提高(如「眼神」)，使得兩者之間的釋義關聯形成偏差。但這樣的現象在所有擴充釋義詞步驟都完成之後(在本例中爲 Lv7)，所有的 DRD 值就會趨於穩定。也就是在教育部所訂的國語辭典中，在「目」字部所有字詞能解釋的範圍下，所有「眼腦」與包括「眼」字雙字詞的釋義關聯都能從圖 5 Lv7 中得到結果。

從圖 5 中亦可看出詞義「眼腦」與「眼睛」的概念相似。雙字詞的多階層 DRD 值較高者，表示則其概念比較明確，而且「眼睛」的語意概念占的比例較高，如「眼語」、「眼科」、「眼罩」等等。多階層 DRD 值在中程度的雙字詞相較於「眼睛」來說，一部分著重在"器官"的概念上，如「眼眶」(0.5)、「眼孔」(0.5)等；另一部分雙字詞的概念則是概念延伸自"目力"、"眼力"，如「眼格」(0.5)、「眼辨」(0.3)等。

多階層 DRD 值小的雙字詞則是含蓋「眼睛」與其它不同的概念於一詞裡，如「眼庫」雖然與「眼腦」共同使用了"外物"、"器官"、"動物"、"身"、"觀察"、"視覺"、"眼睛"七個釋義詞，但「眼庫」釋義原文"國際獅子會中華民國總會爲推動眼角膜移植，接受病人身後捐獻，以協助有眼病的人恢復視力，於民國五十七年十一月十日正式成立的眼角膜儲備設施。"經過濾後保留的字詞很多，且能經多階層擴充的釋義文字少，從而無法增加釋義文字的權重，因而降低了 DRD 值。

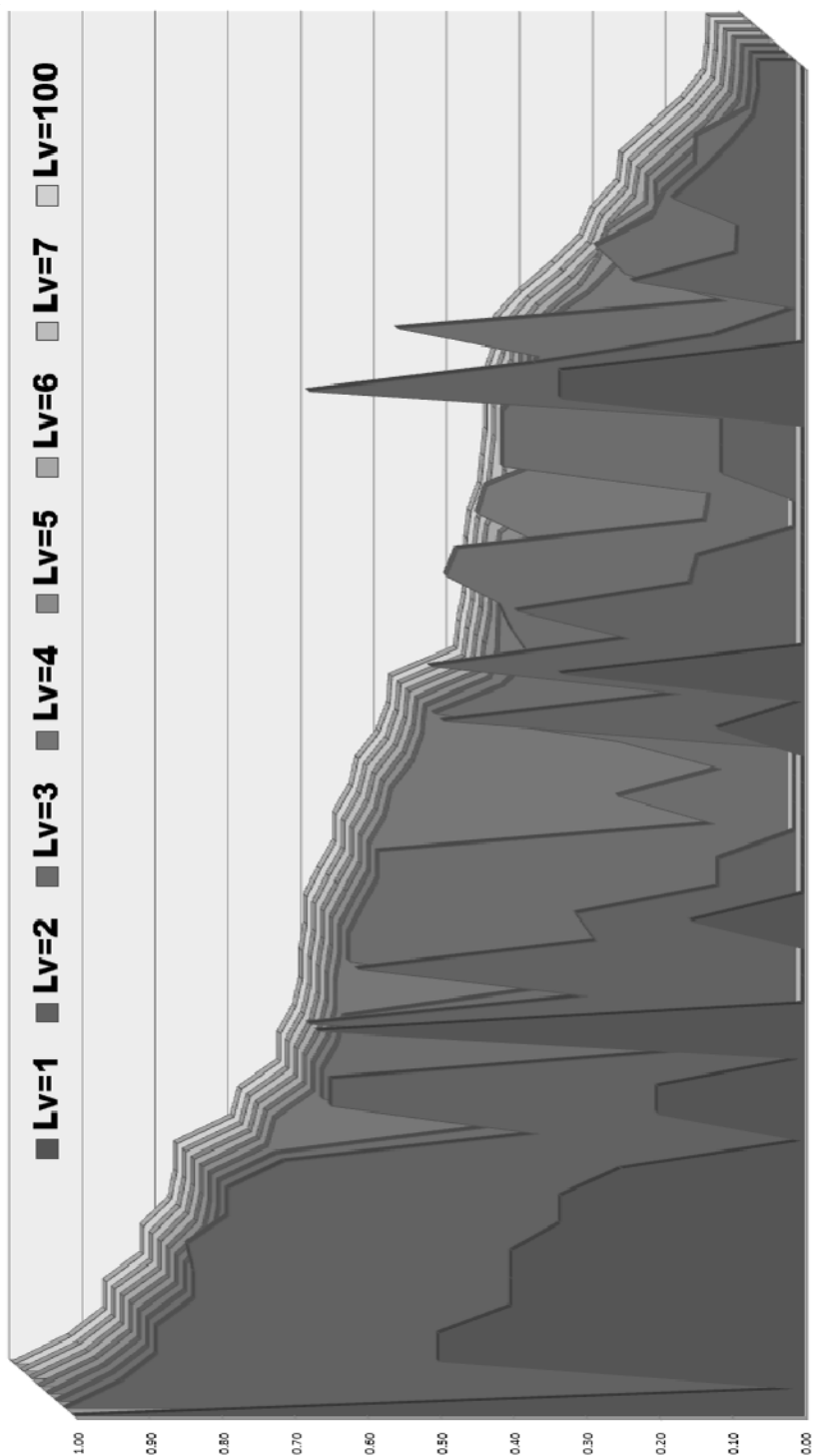


圖5. 包括「眼」字的雙字詞對於「眼腦」之多階層義關聯值比較(Lv1-7, 與Lv100)

爲"目有所恨而止也"。「睪」爲"目視也。从橫目，从幸"。「瞿」爲"鷹隼之視也"。「盾」與「馱」爲"所以扞身蔽目"。上述字若在釋義之中納入《說文解字》的釋義，則 DRD 值則不爲 0。(3)無法判定爲與「目」有關聯的字。「睬」字的釋義"理會"，爲聲符字（即「睬」的聲符是「采」，形符是「目」），其義源於「徇睬」（同「瞅睬」、「徇采」）爲"理睬"。上述除了「睬」字較爲特殊之外，其它部首屬「目」的字都能透過釋義與「目」產生關聯。多階層釋義關聯無法爲 33 個字建立釋義關聯，是因爲本研究中沒有收集部首屬「見」的字詞釋義與《說文解字》之中的釋義。雖然使用於計算的辭典釋義有所局限，但透過多階層釋義關聯計算可發現多數屬「目」部的字與「目」字之間存在有字義上的關聯，進一步的分析也可證明屬「目」部的字與屬「見」部的字詞間存在釋義上的釋義關聯。

3.3 多階層釋義關聯程度與其它語料統計值的比較

我們於 3.2 闡釋了多階層釋義關聯程度值的原則、計算方式及應用，接下來將以語料統計方法中，使用大樣本進行計算的 MI Value 與 t-score 爲例進行比較。以 Mutual Information 計算得到的 MI Value 是在估計兩個字之間的關係(associations)(Church & Hanks, 1990)，MI 的計算爲：

$$I(x, y) \equiv \log_2 \frac{P(x, y)}{P(x)P(y)} \quad (6)$$

其中， x, y 爲兩個待測的字詞，其使用的機率爲 $P(x)$ 、 $P(y)$ ， $I(x, y)$ 表兩者相互共有的訊息。 $I(x, y)$ 在計算上是以聯合機率(joint probability)的方式計算兩個待測字詞之間的獨立觀測機率。透過上述的 MI Value 計算，得到的數值決定兩字詞之間的關聯：(1)若大於 3，則具顯著關聯(Hunston, 2002)；(2)若接近 0 則無顯著關聯；(3)若小於 0，則爲互補關聯。

另一方面，t-score 則是爲了計算兩個詞之間是否具統計上的顯著(Gao & Somers, 1998)，其計算方式爲：

$$t \approx \frac{f(x, y) - \frac{f(x)f(y)}{N}}{\sqrt{f(x, y)}} \quad (7)$$

上述兩個待測的字詞 x, y ，其個別出現的機率爲 $f(x)$ 、 $f(y)$ ， $f(x, y)$ 則爲兩個詞共同出現的次數；而 N 則是所有出現的字詞數。Hunston(2002)指出，當 t-score 大於 2 時，則視此二字具顯著關係。在此我們則以「眼腦」與包括「眼」字雙字詞的關聯計算，來比較多階層釋義關聯、MI Value 與 t-score 值三者的差異(圖 6)。在圖 6 中，y 軸爲計算所得的值。我們可以看到 MI Value 與 t-score 兩者趨勢是與 DRD 值相反，且都呈現正相關。從字義上來看，「眼腦」使用了「眼睛」的釋義；但在 MI Value 與 t-score 值上卻比「眼腦」與「眼蟲」的關聯較高，也就表示 MI Value 與 t-score 在比較兩字詞的關聯度時是不適用的。

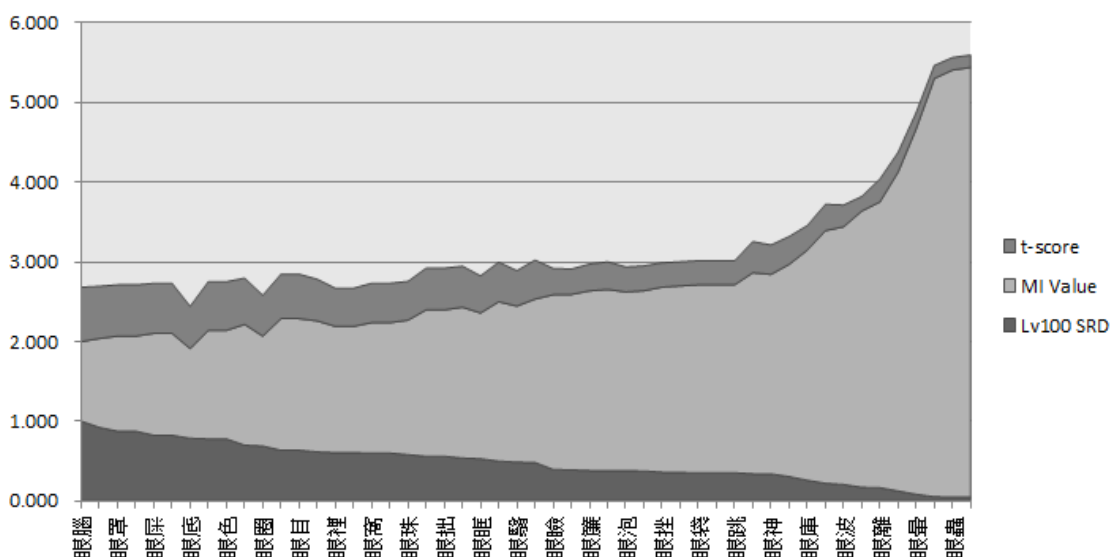


圖6. 「眼腦」與包含「眼」雙字詞之多階層釋義關聯、MI Value 與 t-score 比較

從上述的 MI 值與 t-score 的計算原義來看，二者雖然都是區分兩個字詞在語料庫之中同時出現的頻率的機率比較，但並非是從詞義概念的角度計算兩個字詞之間的關聯。兩者的計算原理與多階層釋義關聯的共用詞所占有的比率雖然相同，但詮釋上卻不同：MI 值是從資訊熵(information entropy)區分兩個字詞所含的釋義字詞是否有共同資訊熵；而 t-score 則是從統計上的顯著性，來區分兩個詞之間的差異。本文提出的多階層釋義關聯的計算是以辭典字詞的釋義做為語料庫的計算基礎，不同於過去 MI 值與 t-score 計算上使用文本(texts)為語料庫基礎的計算。多階層釋義關聯計算的原則不僅可以透過語料庫的內容進行多階層的釋義與擴充外，擴充的字詞的概念亦可從辭典的收集來控制釋義的概念延伸。如前述 3.2 透過同屬「目」的字詞擴展後，計算所有部首屬「目」的字與「目」字之間的關聯。

4. 結論

漢語的知識交流單位為字詞，而且字詞組合的釋義文字亦有知識概念的義涵。本論文試延伸概念隱喻關聯計算至中文的字詞釋義關聯計算，透過辭典對條目的解釋說明文字，提出一套計算詞語釋義關聯的原則。不同於人工的語感詞意分類與語料庫工具計算方式，本文將辭典中的釋義字詞視為計算基準，並進行多階層次擴充以加強釋義字詞的權重，最後在不同的關聯程度之下說明其集群應用與語言上的義涵。本文實作的語料庫來源「教育部重編國語辭典修訂本」，是提供大眾查尋字詞的工具。雖然此辭典自民國八十二年編輯完成後，即很少增改其中內容，但從中取得字詞釋義語料進行計算的結果亦能表現出多階層釋義關聯計算的特色。在本篇文章中，我們討論了單一階層的釋義關聯計算，將與「眼」相關及屬「目」的所有字詞進行計算後，結合社會網路分析工具分析高度釋義關聯集群，不同的集群會依其釋義中共同使用的字詞進行群聚，即同義詞的群聚。

之後我們分析中度的釋義關聯計算中，思考如何處理釋義詞裡與不同概念的權重如何訂定，從而發展出多階層釋義關聯的原則，並說明多階層的釋義對釋義關聯值的變化。接著則分析所有部首屬「目」的字與「目」字之間的多階層釋義關聯，以詮釋漢字義符的特色。並從進一步的分析中得到「目」部字集與「見」部字集的關聯，以及「教育部重編國語辭典修訂本」內的文字釋義裡，沒有特別將字的來源加入，了解多階層釋義關聯在分析漢字與部首之間的使用方式。在本文的最後我們又比較了多階層釋義關聯與 MI Value、t-score 之間的差異，說明了多階層 DRD 值是從概念層級計算兩個字詞之間的交集程度。

本文所介紹的多階層釋義關聯雖然能表現出兩個字詞共同使用的釋義字詞交集程度，但因釋義詞來源為辭典且經過濾而得，因此仍無法完全表達該字或詞的完整概念涵意。此外本研究進行的多階層釋義擴充僅保留動詞與名詞的釋義，且受限於所取得的辭典語料中部首屬「目」的字與詞彙。雖然此一限制於擴充時，能維持「目」部字的字義概念發展方向相同，但仍無法完整擴充所有的字詞。並且在辭典編寫釋義的過程中，若作者沒有將符合條目的所有涵義都編入釋義中，使用多階層釋義關聯計算亦無法完全將詞義的概念均列入運算。縱有上述限制，但透過多階層釋義關聯計算，能將前人所制訂的辭典轉換成爲語言計算的工具。也可再進一步與其它計算語言學的方法及網路分析原則結合，以探討漢字字義與詞義之間的概念及釋義集群中的語意關聯。

參考文獻

- Chung, S.F., Chen, C.H., & Chao, F.Y.A. (2009). Building a Database of Related Concepts of Mandarin Metaphors Based on WordNet and SUMO. In *IEEE International Conference on Semantic Computing*, Berkeley, CA, 378-383.
- Church, K., & Hanks, P. (1990). Word Association Norms, Mutual Information, and Lexicography. *Computational Linguistics*, 16(1), 22-29.
- Dong, Z., & Dong, Q. (2003). HowNet-a hybrid language and knowledge resource. In *IEEE International Conference on Natural Language Processing and Knowledge Engineering*, Beijing, China, 650-655.
- Huang, C.R., Chang, R.Y., & Lee, S.B. (2004). Sinica BOW (bilingual ontological wordnet): Integration of bilingual WordNet and SUMO. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, Lisbon, Portugal.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge University Press.
- Miller, G.A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, 28(11), 39-41.
- 周亞民、黃居仁(2005)。漢字意符知識結構的建立。第六屆漢語詞彙語義學研討會論文集。
- 段玉裁(1808)。《說文解字注》。
- 祝清(2009)。語法隱喻視角的漢語詞典名動詞實證研究。《四川：內江師範學院學報》。

- 高照明(2007)。中文詞彙語意資料的整合及擷取：詞彙語意學的觀點。臺北：第十九屆自然語言與語音處理研討會論文集。
- 許慎(100)。《說文解字》。
- 黃居仁(2005)。漢字知識表達的幾個層面：字、詞與詞義關係概論。漢字與全球化國際學術研討會。臺北。
- 黃居仁、陳聖怡、楊雅君(2008)。意符知識系統研究:[五官類] 意符的概念衍生與知識表徵。第九屆漢語詞彙語義學研討會。新加坡。
- 黃居仁、蔡柏生、朱梅欣、何婉如、黃麗婉(2003)。詞義與義面:中文詞彙意義的區辨與操作原則。In *Proceedings of the Fourth Chinese Lexical Semantics Workshop*, 23-25。
- 鄭文泉(2004)。立象盡意 從符號學的角度論楷體漢字的形上學。國立中央大學哲學所博士論文。
- 黎傳緒(2004)。“相”字的解析。北京教育學院學報，18(4)，17-29。
- 薛榕婷(2003)。說文解字人與自然類部首之文化詮釋。淡江大學中國文學研究所碩士論文。
- 羅益民(2007)。詞典內外的釋義系統。中國辭書學會雙語詞典專業委員會第七屆年會。重慶。

